Achieve

A FRAMEWORK TO EVALUATE COGNITIVE COMPLEXITY IN SCIENCE ASSESSMENTS

SEPTEMBER 2019

Background

Assessment is a key lever for educational improvement. Assessments can be used to monitor, signal, and influence science teaching and learning – provided that they are of high quality, reflect the rigor and intent of academic standards, and elicit meaningful student performances. Since the release of *A Framework for K-12 Science Education* and the Next Generation Science Standards (NGSS), assessment systems are fundamentally changing to surface students' use of disciplinary core ideas (DCI), scientific and engineering practices (SEP), and crosscutting concepts (CCC) *together in service of sense-making* about a phenomenon or problem. As states and districts develop new assessment systems, they need support for developing assessments that balance the vision and integrity of multi-dimensional standards with ensuring that they are sensitive to varying levels of student performance. This brief describes a new approach to capturing and communicating the

complexity of summative assessment items and tasks designed for three-dimensional standards that can be used to ensure that **all learners can make their thinking and abilities visible without compromising the rigor and expectations of the standards.**

A Complexity Framework Focused on Sense-Making

This draft framework for evaluating cognitive complexity in science assessments intentionally builds on expectations for student performance provided by the *Framework for K-12 Science Education* and standards like the NGSS. This complexity framework can be used to determine the degree to which an assessment task asks students to engage in sense-making, a cornerstone of NGSS assessments and performance.

The framework that we propose builds on previous work on cognitive demand in instructional tasks **as** well as criteria and processes for determining the alignment for three-dimensional science assessments. Grounded in sense-making as the overarching umbrella, **this work** is intended to help assessment developers and evaluators determine the degree to which the three dimensions contribute to sense-making within individual items and multi-component tasks as a whole. The framework is organized around two essential questions:

1) **To what degree does the task ask students engage in sense-making?**
2) **In what ways does the task ask students to use each dimension in service of sense-making?**

This framework is designed to support:

- Analyzing the cognitive complexity of individual items (stand-alone and those part of a multi-component cluster) within an assessment.
- Analyzing the cognitive complexity of multi-component tasks as a whole.
- Effective assessment design, development, and evaluation processes.
- Analysis of end-of-instruction assessments at the classroom, district, and state levels.

Components of the Framework

The framework is organized as a two-step process that involves 1) the analysis of individual items or tasks (stand-alone or those comprising a multi-component task; Table 1) and 2) the analysis of multi-component tasks holistically (Table 2).

Step 1: Individual Item Evaluations

The complexity framework first asks users to consider in what ways individual items require students to engage in sense-making. The framework focuses on four indicators that can contribute to higher order thinking and analytical skills in different ways within an item (Table 1):

(1) Scenario contributions to complexity

(2) SEP contributions to complexity

(3) DCI [or disciplinary understanding] contributions to sense-making

(4) CCC contributions to complexity

For each indicator, the complexity framework asks designers and evaluators to consider: **"To what degree does students' engagement with this feature contribute to the level of sense-making required by this task? In what ways does students' use of this dimension support sense-making in this task?"**

For each of these indicators, the contribution of that indicator to student sense-making can range from low to high. Designations of "low", "medium" and "high" complexity for each indicator are based on a combination of indicators-specific factors (e.g., the sophistication of students use of the SEP) as well as some cross-indicator factors such as as how the indicator relates to:

- the **uncertainty/ambiguity** presented in the task;
- **degree of transfer** elicited;
- level and nature of **scaffolding/guidance** embedded in the task; and
- the **range of application and connections** to be made.

Step 2: Holistic Task Analysis

Tasks as a whole are analyzed based on the type and level of thinking the complete performance—across all items and prompts taken individually and as a coherent path of student thinking—requires. Task-level judgments are organized into four categories, drawn from previous work and described in Table 2. These categories focus on how student thinking across dimensions and items work together to support sense-making in the task.

Across the task, the complexity framework asks designers and evaluators to consider **"Across the entire task, what level of sense-making are students being asked to demonstrate?"**

TABLE 1: INDIVIDUAL ITEM ANALYSIS

	Scenario	SEP	DCI [Conceptual - disciplinary]	CCC [Conceptual - crosscutting]
High	Addressing a rich and puzzling phenomenon or problem presented with high-degree of uncertainty.	Figuring out a phenomenon or problem using multiple SEPs in service of authentic sense-making.	Non-routine use of domain specific science ideas as part of sense-making.	Selection and use of conceptual understanding of crosscutting ideas is necessary and expands students' thinking.
Medium	Addressing a phenomenon or problem with some level of uncertainty.	Representation of ideas; use of skills that are relatively complex; some close application.	Supported application of science ideas in typical contexts.	Specific crosscutting concept understanding is needed and is used to focus students' thinking.
Low	Addressing routinely encountered or highly simplified scenarios.	Using the mechanics, skills, and specific knowledge associated with practices isolated from sense-making.	Producing previously-learned ideas and conceptual procedures in routine, well-practiced ways.	Crosscutting concepts are implicitly part of the task, but they are not required in service of sense-making.

For a more detailed description of each indicator and level within the framework, see Table A1 in Appendix A.

TABLE 2: HOLISTIC TASK ANALYSIS

Doing Science
• Student-driven with limited to no scaffolding across all three dimensions; students must decide how to engage and execute within the task. • The three dimensions are used together to engage in sense-making to a high degree throughout the task.
High Guided Integration [low degree of guidance]
• Minimal scaffolding embedded in the task items--students are cued and guided to pursue certain lines of thinking, but have to make some decisions about how and what to engage. • A large majority of the task requires a high degree of sense-making, driven by transfer of thinking and/or uncertainty. • Multidimensional with at least two dimensions equally foregrounded and used in service of making sense of phenomena or problems. Both dimensions' use is sophisticated and at grade-level.
Low Guided Integration [high degree of guidance]
• Moderate to high degree of scaffolding is embedded in the task items to support and guide sense-making—while students have to apply ideas and practices, they are often told which ones to engage and supported in using them. • Students are asked to make sense of a phenomenon or problem they can easily understand but are not inherently familiar with, contributing to the sense-making required. • Multidimensional, but one dimension is often heavily foregrounded--one dimension may routinely be engaged at a lower level of sophistication and/or below grade-level.
Scripted
• Students are provided with well-defined set of actions or procedures are used to complete a given task. • An answer can be obtained with simple application and without significant reasoning. • Focused on obtaining an answer from students' previous understanding or from information provided in the task, not sense making in an effort to understand/explain an uncertainty related to a phenomenon or problem.

For a more detailed description of each level within the task-wide analysis framework, including how each indicator described above contributes to each level, see Table A2 in Appendix A.

Using This Framework

This draft framework is designed to enable item-specific and task-wide complexity judgments as needed. It is intended to be be used in conjunction with criteria and processes for determining alignment to three-dimensional standards. While states, districts, and educators may modify the framework and its use based on their specific contexts and needs, there are some grounding principles and design decisions that may help users make appropriate decisions about how and when to embed this framework in assessment design and implementation processes:

1. Each item receives separate judgments for each of the four indicators. This complexity framework intentionally calls for separate judgments to be made about each indicator because they reveal different and equally important aspects of task complexity. Items with high practice complexity and low disciplinary complexity are fundamentally different than those with low practice and high disciplinary complexity with regard to what kind of student thinking is elicited and how that is interpreted. Figure A1 in Appendix A visually represents what this spread could look like within a task or test.

2. No value judgments are attached to complexity levels. While there is often a tendency to assume that higher complexity tasks are inherently "better" than lower complexity tasks, the appropriate complexity of a task is entirely based on the intended placement, use, and interpretation of student performance. For assessments to be useful, it is essential that tasks give all students the opportunity to demonstrate what they know and can do, and lower complexity tasks (as described by this framework) allow students to do so without compromising the integrity and intent of three-dimensional standards.

3. Designed based on *A Framework for K-12 Science Education*, the framework is designed to work flexibly with all new three-dimensional science standards. This framework acknowledges that different states are adopting and implementing standards that are consistent with the vision embedded in *A Framework for K-12 Science Education* but may differ with regard to specific content, language, progressions, etc. This complexity framework is designed to work flexibly with the range of standards being implemented, by focusing on how practices, disciplinary ideas, and crosscutting concepts contribute to sense-making. As users implement this framework, they should consider using documents associated with the NGSS as a guide (e.g., to determine elements of different practices or grade-specified targets), but may also use their own state- or district-specific documentation.

Acknowledgments

Achieve developed version 1 of this complexity framework for science assessment in close collaboration with Dr. Miray Tekkumru-Kisa, whose Task Analysis Guide-Science (TAGS) framework for science instructional tasks provided a strong foundation for this work. Achieve would like to thank **Dr. Tekkumru-Kisa** (Assistant Professor, Florida State University) and **Aneesha Badrinarayan** (Director, Special Projects and Initiatives, Achieve) for their leadership of this work; state and research partners for their thoughtful feedback and willingness to pilot evolving drafts; and the the Bill & Melinda Gates Foundation, the William and Flora Hewlett Foundation, the Charles and Lynn Schusterman Family Foundation, and Chevron for their generous support of this work. We look forward to updating version 1 based on feedback to refine this approach.

APPENDIX A: DETAILED FRAMEWORK

For those operationalizing this framework, a more detailed description of each indicator and level is included here. An optional numerical scale (1-5) is included for two reasons:

1. Users may find the additional nuance afforded by this scale to be helpful in fostering conversations about intended and targeted task complexity; and
2. Those operationalizing this framework within large scale assessment contexts may find a numerical approach to representing complexity is more conducive to representing and communicating the range of complexity across assessment instruments.

TABLE A1: DETAILED INDIVIDUAL ITEM ANALYSIS RUBRIC

"In what ways does students' understanding and use of [indicator/dimension] contribute to sense-making?"

	Scenario	SEP	DCI [Conceptual - disciplinary]	CCC [Conceptual - crosscutting]
High: 4 or 5	**Emphasis** Addressing a rich and puzzling phenomenon or problem presented with high-degree of uncertainty. • The scenario presents a new phenomenon or problem that is not immediately explainable by student at a level that "figuring out" would be real and authentic for students. This often involves multiple appropriate ways to engage and pursue the task.	**Emphasis** Figuring out a phenomenon or problem using multiple SEPs in service of authentic sense-making. • Addressing the task requires students to engage with grade-specific SEP elements in unexpected, unconventional, or unfamiliar ways in service of sense-making. This may involve the use of multiple SEPs that are not routinely combined, with limited scaffolding. • High degree of student agency in the selection and use of SEPs in ambiguous situations with high-degrees of uncertainty experienced.	**Emphasis** Non-routine use of domain specific science ideas as part of sense-making. • Addressing the task requires students to use and engage in non-typical reasoning with multiple grade-appropriate science ideas. • Ideas are used in service of sense-making. This may involve limited to no scaffolding and far transfer. • High degree of student agency is needed in selection and use of science ideas [content needed is variable or not immediately obvious].	**Emphasis** Selection and use of conceptual understanding of crosscutting ideas is necessary and expands students' thinking. • Addressing the task requires students to engage in complex sense-making that leverages the CCCs. • High degree of student agency is needed in selection, use, and application of the CCCs. • CCCs are used to expand students' thinking into non-typical ways in service of sense-making.
Medium: 3	**Emphasis** Addressing a phenomenon or problem with some level of uncertainty. • The scenario presents a relatively new phenomenon that students might have some familiarity with, but do not fully understand the specific uncertainty the task is focused on. • The task provides a scenario with multiple facets of information for students to interpret at a grade-appropriate level of sophistication • Students are provided with some explicit cues and/or scaffolding to engage in the scenario,.	**Emphasis** Representation of ideas; use of skills that are relatively complex; some close application. • Addressing the question requires students to engage in grade-specific SEPs in expected or well-practiced ways to demonstrate the use of previously developed ideas or to engage in routine sense-making. This may involve the use of multiple SEPs that are routinely coupled, with some scaffolding. • Close application to familiar/expected contexts (i.e., near transfer) may be needed. • Students are required to demonstrate some understanding of how/why to use the SEP.	**Emphasis** Supported application of science ideas in typical contexts. • Addressing the task requires students to use grade-appropriate science ideas as part of student reasoning in typical contexts with routine, well-practiced ways. • Addressing the task may require students to connect multiple ideas in routine ways. • The task may include some scaffolding and cuing about which ideas to use but still requires students to apply their understanding of the science ideas.	**Emphasis** Specific crosscutting concept understanding is needed and is used to focus students' thinking. • Addressing the task requires students to engage in grade-specified CCCs in scaffolded/cued ways to focus students' thinking.
Low: 1 or 2	**Emphasis** Addressing routinely encountered or highly simplified scenarios. • The task provides a problem or a phenomenon that students are already familiar with how to explain or solve.	**Emphasis** Using the mechanics, skills, and specific knowledge associated with practices isolated from sense-making. • Addressing the task requires students to demonstrate simple, procedural, and mechanical aspects of engaging in SEPs (reading graphs/charts, drawing diagrams, etc.). • Students may be provided with a script/set of defined procedures to follow to engage with the SEP, with limited student thinking required about which, how, or why practices are engaged. • Practice demand is below grade level.	**Emphasis** Producing previously-learned ideas and conceptual procedures in routine, well-practiced ways. • Addressing the task requires direct representation of previously learned grade-level ideas and concepts, including well-developed procedures related to concepts (e.g., Punnett squares). • Addressing the task does not require relating science ideas to one another, reasoning with ideas or using them in service of sense-making.	**Emphasis** Crosscutting concepts are implicitly engaged or practiced, but are not consciously used by students in service of sense-making. • Addressing the task inherently involves the CCCs (e.g., explanation involving a cause and effect) but does not require students to demonstrate that they understand and can use elements of the CCCs. • CCCs which are employed in the task are below grade level.

TABLE 2: HOLISTIC TASK ANALYSIS

In this table, we include additional guidance for those making holistic task judgments, including how the individual components of the framework may contribute to the holistic judgment.

Doing Science	
• Student-driven with limited to no scaffolding across all three dimensions; students must decide how to engage and execute within the task. • The three dimensions are used together to engage in sense-making.	**Scenario:** A phenomenon-based scenario is presented with uncertainty/ambiguity at a level that "figuring out" would be real and authentic for students without a clear pathway to follow. **SEP**: Students work like scientists to use various scientific practices to be able to develop or deepen an understanding of a scientific idea or problem as they explore a phenomenon. **DCI:** Students engage in a high degree of decision making regarding which ideas to employ and how to employ them, often using sophisticated reasoning with multiple dimensions. **CCC:** Multiple CCCs are used to expand student sense-making.
High-Guided Integration [Low guidance]	
• Minimal scaffolding embedded in the task items--students are cued and guided to pursue certain lines of thinking, but have to make decisions about how and what to engage. • A large majority of the task requires a high degree of sense-making, driven by transfer of thinking and/or uncertainty. • Multidimensional with at least two dimensions equally foregrounded and used in service of making sense of phenomena or problems. Both dimensions' use is sophisticated and at grade-level.	**Scenario**: Students are required to explain a rich and puzzling phenomenon or address a problem by connecting what they have learned and experienced in a new way (i.e., far transfer) with guidance for how to engage, but leaving many decisions to the student. **SEP:** Students use, with scaffolding, the SEPs to connect and negotiate multiple variables, factors, perspectives, etc. to engage in sense-making/reasoning processes employed in real scientific inquiry. **DCI:** Students are required to employ complex (possibly non-typical), sophisticated reasoning using science ideas. Students may be connecting multiple complex, distinct, but distally related ideas with guidance to help students navigate pulling these ideas together. **CCC:** CCCs are used to expand student sense-making.
Low-Guided Integration [High guidance]	
• Moderate to high degree of scaffolding is embedded in the task items to support and guide sense-making—while students have to apply ideas and practices, they are often told which ones to engage and supported in using them. • Students are asked to make sense of a phenomenon or problem they can easily understand but are not inherently familiar with, contributing to the sense-making required. • Multidimensional, but one dimension is often heavily foregrounded--one dimension may routinely be engaged at a lower level of sophistication and/or below grade-level.	**Scenario-**:Students are asked to make sense of a simple but unique problem or phenomenon, with multiple facets of information for students to interpret, and students are guided in that interpretation. The phenomenon is not immediately explainable by students recall or restatement of science ideas and practices (i.e., not simply confirmatory application of the DCI) **SEP:** Emphasis on using SEPs to represent ideas with some application and reasoning. **DCI:** Some sense-making by connecting content to a phenomenon is needed to successfully complete the question. The focus is on students' demonstrating their understanding of content through application by reasoning with the science ideas in typical/expected ways. Students might be prompted or guided to focus their thinking on targeted science ideas, but they need to decide how to use them to respond to the question. **CCC:** Implicit OR specific targeted understanding, CCC is used to focus student thinking.
Scripted	
• Students are provided with well-defined set of actions or procedures are used to complete a given task. • An answer can be obtained with simple application and without significant reasoning. • Focused on obtaining an answer from students' previous understanding or from information provided in the task, not sense making in an effort to understand/explain an uncertainty related to a phenomenon or problem.	**Scenario:-** Students are required to explain a phenomenon or address a problem that students can be expected to fully understand (e.g., Students were likely asked to address the same (or extremely similar) phenomena or problem; phenomena/problem does not present uncertainty. The scenario includes limited information for students to interpret; if multiple modalities are used, they are used to convey the same information. **SEP:** Focus on using the mechanics of SEPs without requiring students to think about the reasons for when and why these procedures or skills are needed (e.g., with heavy scaffolding). **DCI:** The emphasis of the content is to provide an opportunity for students to demonstrate that they "know" definitions, concepts with very direct/contrived application. Students are often told what to do and which ideas to represent within the task items. **CCC:** mplicit use of CCCs.

FIGURE A2: VISUALIZING THE DISTRIBUTION OF ITEMS ACROSS INDICATORS

One of the principles guiding the development of this framework is that each indicator described contributes to sense-making and student performance in different and important ways—and as a result, it is critical that we capture and use those differences when connecting student performance on an assessment with feedback, proficiency determinations, and next steps or interventions. By attending to specific distributions, we can see trends and gaps in the kind of thinking elicited, and incorporate these features into interpretation and next steps to promote student sense-making.